



November 21, 2024

VIA ECF

Honorable Ona T. Wang
United States District Court
500 Pearl Street
New York, NY 10007

Re: *The New York Times Company v. Microsoft Corp., et al.*, No. 23-cv-11195-SHS; *Daily News, L P, et al. v. Microsoft Corp., et al.*, 1:24-cv-3285; *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, 1:24-cv-04872: Response to Motion to Compel Early Deposition on Microsoft's Log Data

Dear Magistrate Judge Wang:

Microsoft submits this response to Plaintiffs Daily News, LP et al., The New York Times Company, and The Center for Investigative Reporting, Inc. (collectively "Plaintiffs") November 18, 2024 letter (ECF 311), in which Plaintiffs seek to compel an accelerated "custodial" 30(b)(6) deposition by December 20, 2024, regarding Microsoft's data logs. Microsoft has agreed that a properly tailored 30(b)(6) deposition on complex data could be appropriate, but it is a bad idea to force Plaintiffs' demand for a broad and improper deposition on short notice. Further, Microsoft produced samples and documentation describing the data, but rather than ask a single question, Plaintiffs short circuited the meet and confer process after the October 30 hearing in the guise of seeking an immediate "custodial" deposition. There are three independent reasons why the Court should deny Plaintiffs' request: (1) the proposed deposition regarding voluminous data logs is not a "custodial" deposition; (2) a complex data deposition is premature; and (3) the "topics" served by Plaintiffs fail to meet requirements of Rule 30(b)(6).

What Are The Data Logs At Issue? Early in discovery, Microsoft disclosed ESI in the form of data regarding usage of Copilot. Now more than 15 petabytes, this massive data store consists of more than 300 distinct data fields in different log formats. Some of these fields are relevant to Plaintiffs' claims, but the vast majority are not (e.g., data re: latency of responses). The data is difficult to access and cannot be reproduced or directly inspected. Different data fields are used by distinct teams within Microsoft for distinct purposes, such that no one person has knowledge of all the logs and data fields. Rather, persons knowledgeable about all the different logs and fields are situated globally.

Procedural Background. Microsoft *voluntarily* collected and produced documentation regarding the data fields and logs, and also generated and produced a sample showing exemplar data associated with each of the hundreds of field types. The parties were in the midst of discussing next steps just prior to the October 30, 2024 hearing.

At that hearing, the Court asked the parties to consider whether an early 30(b)(6) deposition on *custodial issues* would help to resolve *custodial discovery disputes*. See 10/30/24 Tr. at 8:15-20, 62:21-23. Plaintiffs seized upon the Court's comment to try to force an immediate deposition regarding the user data logs, identifying two subjects related to such logs as part of their demand:

Honorable Ona T. Wang

November 21, 2024

15. The existence, organization, preservation, storage, and deletion of output data reflecting user sessions on Your Generative AI Products and Services, as well as of reports on, metrics on, or analysis of such output data; and

16. The meaning of data fields contained in sample output logs produced by Microsoft at MSFT_NEWS_000624874, MSFT_NEWS_000624875, and MSFT_NEWS_000624876, as well as of data fields contained in other output logs maintained by Defendants.” Ex. 1.

Microsoft objected, noting that these are not custodial deposition subjects. Nevertheless, Microsoft indicated that it would be willing to offer a witness on properly drawn topics in a reasonable time frame. Microsoft asked Plaintiffs to provide the draft of the actual 30(b)(6) Notice that they proposed to conduct and to meet and confer about scope and timing. Plaintiffs refused, responding that these numbered paragraphs in fact constituted their proposed deposition notice. ECF 311-1. Ex. 1. Plaintiffs also insisted that such a deposition be conducted before year end.

A Data Log Deposition Is Not Custodial Discovery. The purpose of a custodial deposition, namely discovery about discovery, is to ascertain whether a party has conducted an adequate collection. *See Trustees of Local 854 Pension Fund v. Barrett*, No. 23 Civ. 1160, 2024 WL 4026257, at *5 (S.D.N.Y. Sept. 3, 2024); *Winfield v. City of New York*, No. 15-cv-05236, 2018 WL 840085, at *3 (S.D.N.Y. Feb. 12, 2018) (Party seeking discovery on discovery must provide an adequate factual basis to justify the discovery, and the Court must closely scrutinize the request). As described in Microsoft’s portion of the November 1 joint submission (ECF 306), there is no basis for such a deposition of Microsoft.

In any event, a custodial deposition would not encompass Plaintiffs’ broad topics. The data logs are in no way “custodial.” They are not in the possession of any individual custodian nor is there concern about whether they have been “collected.” These topics are not designed to ascertain whether Microsoft has complied with its discovery obligations, but rather, are designed to assist Plaintiffs’ experts in devising a querying strategy to inspect the data. A deposition about complex data using properly scoped topics in a reasonable time frame would make sense after the parties engaged in a meaningful informal exchange, not as part of custodial discovery. Plaintiffs refused.

A Data Log Deposition Is Premature. In addition to being improper topics for custodial discovery that the Court asked the parties to consider, rushing a 30(b)(6) deposition about the data logs will not streamline discovery. As Microsoft’s voluntary disclosures demonstrate, this collection comprises more than 15 petabytes of data consisting of over 300 distinct data fields and multiple log formats. Determining how best to analyze and query this amount of data is a complex problem requiring thought, cooperation, and communication between the parties. Indeed, “especially with respect to depositions concerning data, it is the joint responsibility of the parties to cooperate” to facilitate the “exchange of accurate information in a cost-effective and efficient manner.” *Winfield*, 2018 WL 840085, at *6. Plaintiffs’ request for a deposition requiring Microsoft to provide multiple witnesses on the contents of hundreds of data fields, and on reports, metrics, and analyses of the data logs, without first attempting to work with Microsoft to narrow the topics to relevant data fields and reports is neither cost-effective nor efficient. Simply put, Plaintiffs refused to analyze the samples Microsoft provided and to then discuss and tailor a deposition to what they actually need. Instead they rushed to put this manufactured dispute before the Court.

Winfield is on point. There, plaintiffs sought a 30(b)(6) deposition regarding a voluminous and complex database. *Id.* at *1 n.2. Before allowing the deposition, this Court first encouraged informal meetings with plaintiffs’ counsel and technology experts to facilitate the exchange of

Honorable Ona T. Wang

November 21, 2024

information about the location, availability, and limitations of the data. *Id.* The same should happen here. Taking a premature deposition on the meaning of over 300 data fields, “as well as [] data fields contained in other [unidentified] output logs maintained by Defendants” is not an efficient method of providing Plaintiffs with the information their experts require. Plaintiffs assert they have been negotiating with Microsoft for months, but that is not quite right. It bears repeating that Microsoft and OpenAI are separate companies, with separate legal teams, and distinct discovery issues. Here, Plaintiffs appear to be citing their conferral with OpenAI regarding inspecting OpenAI’s training data and models. Dkt. 311 at 2. Indeed, two of the exhibits Plaintiffs attach to their motion are requests for inspection to OpenAI, *not* Microsoft. Exs. B, E. In the nearly one month since Microsoft provided documentation and data samples, Plaintiffs have consistently *refused* to work with Microsoft to understand the data or seek information via informal means.

Microsoft requests the Court instruct Plaintiffs to examine the data samples, confer with their experts, and seek additional information informally from Microsoft to better understand the data and associated challenges with querying the data before jumping straight to a deposition on every data field. Doing so will allow Plaintiffs to propound tailored and reasonably particular topics about which Microsoft can properly educate witnesses, and will allow Plaintiffs to ask informed questions at the deposition. What Plaintiffs are demanding will waste time and expense and will almost inevitably lead to avoidable disputes about whether witnesses were properly prepared.

The “Topics” Are Not Described With Requisite Particularity. Deposition topics must be described with reasonable particularity. Fed. R. Civ. P. 30(b)(6). Where the topics concern complex data, even greater specificity is required to ensure that the witness(es) can prepare for the deposition and that the deposition is productive. *Winfield*, 2018 WL 840085, at *5 (citing *EEOC v. Thorman & Wright Corp.*, 243 F.R.D. 421, 426 (D. Kan. 2007) (requiring “painstaking specificity”)). In the context of depositions regarding data systems, courts consider: (1) the nature of the topics; (2) whether the descriptions of the topics include examples of questions and clarifying information; (3) whether a reasonable person reading the notice would understand how to prepare for the deposition; (4) the level of specificity required; (5) the number of witnesses needed to respond to the topics; (6) the costs and burden on the corporate party to adequately prepare the witnesses; and (7) the availability of less burdensome or costly methods to obtain the information sought. *Id.* at *6-7.

Topics 15 and 16 do not meet this standard. Contrary to Plaintiffs’ assertion that these are “simpl[e]” topics (ECF 311-1 at 2), they are extremely broad and vague, requiring a witness to be knowledgeable regarding (1) the meaning of hundreds of data fields, (2) unspecified reports and metrics about the data, and (3) the meaning of data fields “contained in other [unspecified] output logs.” It would be impossible for Microsoft to educate a single witness on these broad subjects. *See Beverley v. New York City Health & Hosps. Corp.*, No. 18-CV-08486 (ER), 2024 WL 2125402, at *8 (S.D.N.Y. May 13, 2024) (topics requiring memorizing an unrealistic amount of information found overbroad and not described with reasonable particularity). Further, a global team is knowledgeable about the user data logs, some of whom do not speak English as a first language. Specificity of topics is required so that Microsoft can identify the right personnel and obtain the relevant information to educate the corporate witness. These overbroad topics make that impossible and would most likely result in a time-wasting exercise in futility.

Winfield is again on point. This Court refused a deposition much like the one sought here and instead required the plaintiffs’ experts to analyze the produced data and provide specific questions ahead of the deposition to allow the defendant to adequately prepare. *Id.* at *8. The Court should do the same here.

Honorable Ona T. Wang

November 21, 2024

Respectfully submitted,

/s/ Annette L. Hurst

Annette L. Hurst

Counsel for Defendant Microsoft Corporation

Respectfully submitted,

/s/ Jared B. Briant

Jared B. Briant

cc: Counsel of Record (via ECF)